CLAIMS:

1. A document categorizing method for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said method being characterized in that:

after categorizing said plurality of documents into a plurality of clusters according to semantic similarity, a cluster merging process is performed such that relations among clusters of said plurality of clusters are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together.

2. A document categorizing method according to Claim 1, wherein said cluster merging process is performed such that the evaluation of relations among clusters under consideration as to whether they should be merged or not is performed on the basis of the number of documents commonly included in said clusters under consideration relative to the total number of documents included in said clusters under consideration, and cluster merging is performed in accordance with the evaluation result.

3. A document categorizing method according to Claim 1, wherein said cluster merging process is performed such that in what manner feature elements, which characterize respective clusters under consideration as to whether they should be merged or not, appear in the respective clusters under consideration is examined, and cluster merging is performed in accordance with the manner in which the feature elements appear.

4. A document categorizing method according to one of Claims 1 to 3, wherein said cluster merging process is performed at least for two clusters, and after completion of the cluster merging process a first time, the cluster merging process is performed repeatedly for the resultant set of clusters until no further cluster merging occurs.

5. A document categorizing method according to one of Claims 1 to 4, wherein after completion of said cluster merging process, supplementary information indicating that cluster merging has been performed and also indicating the basis on which the cluster merging has been performed is output.

6. A document categorizing method for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said method being characterized in that:

after categorizing said plurality of documents into a plurality of clusters according to semantic similarity, a cluster merging process is performed such that relations among clusters of said plurality of clusters are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together; and

information representing which clusters have bee merged together and also representing the degrees of relation among the merged clusters is generated and said information is output together with the categorization result to be presented to a user so that when final clusters obtained as a result of said cluster merging process are displayed, the user can see in what manner said cluster merging process has been performed to obtain said final cluster.

7. A document categorizing method according to Claim 6, wherein said information output so as to enable the user to see in what manner said cluster merging process has been performed is given by modifying the manner of displaying the cluster names of respective clusters merged together in accordance with the degree of relation among said clusters merged together in such a manner that when said degree of relation among said clusters is higher than a predetermined value, said cluster names are displayed in an AND form, however when said degree of relation among said clusters is lower than the predetermined value, said cluster names are displayed in an OR form.

8. A document categorizing method according to Claim 7, wherein when said cluster names are displayed in the AND form, said cluster names of the respective clusters are displayed successively in a single horizontal line or the respective cluster names are displayed in different lines, while when said cluster names are displayed in the OR form, a delimiter is inserted between adjacent cluster names of the respective clusters.

9. A document categorizing method according to Claim 7 or 8, wherein when a certain cluster includes a cluster therein, the name of said cluster included in said certain cluster is enclosed within brackets and placed after the name of said certain cluster.

10. A document categorizing apparatus for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said apparatus comprising:

a clustering unit for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity; and

a cluster merging unit which evaluates the relation among the plurality of clusters created by the clustering unit on the basis of the documents included in the respective clusters and then combines two or more clusters having a degree of relation equal to or higher than a predetermined value.

11. A document categorizing apparatus for categorizing a plurality of documents into a plurality of clusters according to semantic similarity, said apparatus comprising:

a clustering unit for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity,

a cluster merging unit which evaluates the relation among the plurality of clusters created by the clustering unit on the basis of the documents included in the respective clusters and then combines two or more clusters having a degree of relation equal to or higher than a predetermined value;

a cluster-merging-process information generator for generating cluster-merging-process information representing which clusters have been merged together and also representing the degrees of relation among the merged clusters wherein said cluster-merging-process information is to be displayed when final clusters obtained via said cluster merging process performed by said cluster-merging unit are displayed so that a user can see in what manner said cluster merging process has been performed to obtain said final cluster; and

categorization result outputting means for outputting said cluster-merging-process information such that said cluster-merging-process information is included in the categorization result to be presented to said user.

12. A storage medium on which a document categorizing program for categorizing a plurality of documents into a plurality of clusters according to semantic similarity is stored, said document categorizing program comprising:

a clustering step for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity, and

a cluster merging step in which the degrees of relation among clusters of said plurality of clusters obtained in said clustering step are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a degree of relation equal to or higher than a predetermined value are combined together.

13. A storage medium on which a document categorizing program for categorizing a plurality of documents into a plurality of clusters according to semantic similarity is stored, said document categorizing program comprising:

a clustering step for categorizing a plurality of documents into a plurality of clusters in accordance with semantic similarity;

a cluster merging step in which the degrees of relation among clusters of said plurality of clusters obtained in said clustering step are evaluated on the basis of documents included in the respective clusters, and two or more clusters having a

degree of relation equal to or higher than a predetermined value are combined together;

a cluster-merging-process information generating step for generating cluster-merging-process information representing which clusters have been merged together and also representing the degrees of relation among the merged clusters wherein said cluster-merging-process information is to be displayed when final clusters obtained via said cluster merging process performed by said cluster merging unit are displayed so that a user can see in what manner said cluster merging process has been performed to obtain said final cluster; and

a step for outputting said cluster-merging-process information such that said cluster-merging-process information is included in the categorization result to be presented to said user.